Computational Design of a Protein Crystal

Christopher J Lanci  
*University of Pennsylvania*

Christopher M MacDermaid  
*University of Pennsylvania*

Seung-gu Keng  
*University of Pennsylvania*

Rudresh Acharya

Benjamin North

See next page for additional authors

Follow this and additional works at: https://repository.upenn.edu/chemistry_papers

Part of the Biochemistry Commons, Organic Chemistry Commons, and the Physical Chemistry Commons

**Recommended Citation**


This paper is posted at ScholarlyCommons. https://repository.upenn.edu/chemistry_papers/22
For more information, please contact repository@pobox.upenn.edu.
Computational Design of a Protein Crystal

Abstract
Protein crystals have catalytic and materials applications and are central to efforts in structural biology and therapeutic development. Designing predetermined crystal structures can be subtle given the complexity of proteins and the noncovalent interactions that govern crystallization. De novo protein design provides an approach to engineer highly complex nanoscale molecular structures, and often the positions of atoms can be programmed with sub-Å precision. Herein, a computational approach is presented for the design of proteins that self-assemble in three dimensions to yield macroscopic crystals. A three-helix coiled-coil protein is designed de novo to form a polar, layered, three-dimensional crystal having the P6 space group, which has a “honeycomb-like” structure and hexameric channels that span the crystal. The approach involves: (i) creating an ensemble of crystalline structures consistent with the targeted symmetry; (ii) characterizing this ensemble to identify “designable” structures from minima in the sequence-structure energy landscape and designing sequences for these structures; (iii) experimentally characterizing candidate proteins. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Å agreement [backbone root mean square deviation (rmsd)] with the computational model of the crystal. This approach to crystal design has potential applications to the de novo design of nanostructured materials and to the modification of natural proteins to facilitate X-ray crystallographic analysis.

Keywords
biomaterials, computational protein design, crystal engineering, protein crystallization, self-assembly

Disciplines
Biochemistry | Organic Chemistry | Physical Chemistry

Author(s)
Christopher J Lanci, Christopher M MacDermaid, Seung-gu Keng, Rudresh Acharya, Benjamin North, Xi Yang, X Jade Qiu, William F. DeGrado, and Jeffery G. Saven

This journal article is available at ScholarlyCommons: https://repository.upenn.edu/chemistry_papers/22
Protein crystals have catalytic and materials applications and are central to efforts in structural biology and therapeutic development. Designing predetermined crystal structures can be subtle given the complexity of proteins and the noncovalent interactions that govern crystallization. De novo protein design provides an approach to engineer highly complex nanoscale molecular structures, and often the positions of atoms can be programmed with sub-Å precision. Herein, a computational approach is presented for the design of proteins that self-assemble in three dimensions to yield macroscopic crystals. A three-helix coiled-coil protein is designed de novo to form a polar, layered, three-dimensional crystal having the P6 space group, which has a “honeycomb-like” structure and hexameric channels that span the crystal. The approach involves: (i) creating an ensemble of crystalline structures consistent with the targeted symmetry; (ii) characterizing this ensemble to identify “designable” structures from minima in the sequence-structure energy landscape and designing sequences for these structures; (iii) experimentally characterizing candidate proteins. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Å agreement [backbone root mean square deviation (rmsd)] with the computational model of the crystal. This approach to crystal design has potential applications to the de novo design of nanostructured materials and to the modification of natural proteins to facilitate X-ray crystallographic analysis.

Computational design of a protein crystal

Christopher J. Lanci,1,2 Christopher M. MacDermaid,1,3 Seung-gu Kang,1 Rudresh Acharya,2 Benjamin North,1 Xi Yang,1X. Jade Qiub,c, William F. DeGradob,c, and Jeffery G. Savena,2

1Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-6232; 2Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6059; and 3Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158-9001

Edited by David Eisenberg, University of California, Los Angeles, CA, and approved March 8, 2012 (received for review August 2, 2011)

Protein crystals have catalytic and materials applications and are central to efforts in structural biology and therapeutic development. Designing predetermined crystal structures can be subtle given the complexity of proteins and the noncovalent interactions that govern crystallization. De novo protein design provides an approach to engineer highly complex nanoscale molecular structures, and often the positions of atoms can be programmed with sub-Å precision. Herein, a computational approach is presented for the design of proteins that self-assemble in three dimensions to yield macroscopic crystals. A three-helix coiled-coil protein is designed de novo to form a polar, layered, three-dimensional crystal having the P6 space group, which has a “honeycomb-like” structure and hexameric channels that span the crystal. The approach involves: (i) creating an ensemble of crystalline structures consistent with the targeted symmetry; (ii) characterizing this ensemble to identify “designable” structures from minima in the sequence-structure energy landscape and designing sequences for these structures; (iii) experimentally characterizing candidate proteins. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Å agreement [backbone root mean square deviation (rmsd)] with the computational model of the crystal. This approach to crystal design has potential applications to the de novo design of nanostructured materials and to the modification of natural proteins to facilitate X-ray crystallographic analysis.
The target protein was a mathematically created (idealized) homotrimeric parallel-coiled-coil protein having C3 symmetry, a superhelical pitch of 120 Å, and (initially) 27 residues per helix (36). Similar structures have been observed in prototypical three-stranded coiled coils (37–39). The 8 interior (a and d heptad) positions of each helix contained well-packed Val and Leu residues (39). At the remaining exterior positions, all amino acids except cysteine and proline were considered in the design of crystalline ordering.

Within a crystal of P6 symmetry, many symmetry-related configurations involving the multiple copies of the protein are possible. Choosing a high-symmetry space group reduces the number of degrees of freedom D available to the protein. In general, protein units within P6 have D = 5 (40), but here D is reduced to D = 2 when generating the crystalline configurations: (i) The C3 axis of the protein is chosen to coincide with a C3 axis in the crystal (Fig. 1A), and (ii) the length of the c-unit cell vector is fixed at c = 40.7 Å to achieve backbone helical hydrogen bonding between helix termini (Fig. 1B). Each helix’s approximately 28 residue equivalents displace the N terminus approximately 120° (one-third turn) about the protein’s superhelical C3 axis relative to the C terminus, resulting in a pseudointercontiguous super helix and hydrophobic core. Complementary hydrogen bonding and hydrophobic interactions are specified at the interlayer interface. Candidate crystalline arrays consistent with P6 were generated by varying intralayer degrees of freedom R and θ. R is the distance between neighboring proteins (a = b = 31/2R), θ is the angle of rotation of each protein about its superhelical C3 axis (Fig. 1A), and this rotation maintains the C32 axis between nearest neighbors. N-terminal acetylated proteins were used to characterize crystalline arrays of structures energetically. All sequence design calculations were performed in the context of the local 3-D crystalline array.

In a first attempt, a minimum energy crystalline configuration was identified based upon backbone interactions on the protein. “Minimal side chain” calculations were used where Gly was modeled at each of the protein’s exterior positions. A grid search was performed over R and θ (15 Å < R < 22 Å, increment ∆R = 0.2 Å; 0 < θ < 60°, increment ∆θ = 5°, 520 structures), which identified the lowest energy configuration (R = 19.0 Å, θ = 0.0°) [AMBER (41)]. Subsequently, a sequence was computationally designed using this structure. In the local crystalline environment (42), the exterior residues of each helix (19/27) were computationally designed consistent with the periodic symmetry (42, 43).

**Experimental Characterization of Protein P6-a.** A single low-energy sequence for the identified structure, P6-a, was selected, synthesized and formed diffusion-quality crystals overnight at room temperature using a standard crystal screen. P6-a crystallized in the apolar P321 space group, and the structure was solved to 2.9 Å resolution. The crystal contained columnar, hexagonal pores resembling the target, but neighboring proteins were antiparallel (Fig. 2A). A model of the P6-a sequence in the observed P321 crystalline array was built that contains the most probable side-chain conformations. For the crystallographic structure and this model, the computed interaction energy per protein within the P321 structure was above that of the P6 structure, i.e., the calculation did not discern P321 as the preferred crystalline structure. The observed frequency of P321, however, is more than three times the frequency of P6 (21, 40). In addition, P321 can accommodate deviations from the planar layering present in P6. The formation of P321 by this protein may be kinetically and entropically more facile than that of P6. Furthermore, the antiparallel packing involves an extended right-handed “glycine zipper” motif (GX2GX4A) (44), which is similar to the GX3G motif but is found in both parallel and antiparallel orientations (45, 46).

**Results and Discussion**

**Computational Design.** In many applications of protein design, desirability of a structure is made accessible by using structures from naturally occurring proteins (24, 27, 32, 33), but here both the structure of the protein as well as its crystalline ordering are specified de novo. A trimeric coiled-coil protein was designed de novo to form a polar, layered three-dimensional crystal having the P6 space group. Candidate proteins were identified from local energy minima on a sequence-structure energy landscape. X-ray crystallographic studies revealed that one such designed protein is in subangstrom (Cα rmsd < 0.70 Å) agreement with the computationally designed model.

*Fig. 1. (A) One layer of the P6 crystal viewed along the unit cell c-axis. Protein (open circle) comprises three identical helices (small filled circles). R is the intralayer distance between neighboring proteins. θ is the rotation angle about the protein’s superhelical axis. C3 symmetric rotation axes (N-sided polygons) and C3 axes (ovals) indicated. (B) Two adjacent layers of the P6 crystal and helical hydrogen bonding at the interlayer interface.*
Sequence-Structure Energy Landscape. Given the subtlety of engineering proteins consistent with the targeted polar crystal, the scope of the design was broadened to characterize a wide range of sequences and structures concomitantly, i.e., to survey the sequence-structure designability landscape of proteins consistent with P6 symmetry. The protein was shortened to 26 residues (28 counting the N-terminal acetyl and the C-terminal amide (CO-NH₂) groups as residues) (39), in order to permit a one- amide (1.5 Å rise/residue) gap between helix termini that ameliorates possible steric interactions and allows facile readjustment of the individual coiled-coil proteins upon crystallization. A higher resolution sampling of the sequence-structure energy landscape was undertaken (15 Å < R < 22 Å, ΔR = 0.1 Å; 0 < θ < 120°, Δθ = 0.5°; 19,200 structures) to allow greater possibility for identifying proteins specific for P6 symmetry. R and θ values, where the backbone atoms overlapped, were discounted. In the context of a protein’s nearest neighbors within each crystalline configuration, the site-specific probabilities of the 18 allowed amino acids, their side-chain conformations, and the weighted average energy over these sequence-rotamer states E(R, θ) were estimated computationally (30, 31, 47). Effective solvation energy functions were not used (31), and no constraint was placed on the net charge of the protein. A symmetry assumption was applied, where equivalent residues on all helices had the same identities and same rotamer states (42). This resulted in variation of only 18 unique, exterior residues. E(R, θ) captures the contributions of predominantly low-energy sequences (31, 48), and it yields a sequence-structure energy landscape (Fig. 3). Minima identify candidate crystalline structures that potentially support such low-energy sequences. In arriving at specific sequences for given R and θ, the calculations proceeded iteratively. With each iteration, the most probable amino acid was selected at sites where its probability was at least twice the next most probable. In the final iteration, a helix propensity (49) constraint was imposed and chosen to have a value consistent with parallel homotrimeric proteins in the crystal. A pair of laterally neighboring proteins (P6-d) comprising only helix D. The two protein structures are nearly identical. The GX₇G motif is displaced one heptad nearer the C terminus in P6-d than P6-a. Interior residues are gray; “ace” denotes acetylated N terminus.

Comparison Between the Model and Crystallographic Structures. The P6-d proteins are oriented in the targeted all-parallel polar arrangement with R = 18.4 Å, in agreement with the template model (R = 18.8 Å). Interestingly, the crystal structure of protein P6-d has four helices (ABCD) in the asymmetric unit, which differ mostly in side-chain conformations, yielding two distinct protein structures: P6-d.1, comprising helices A, B, and C within the asymmetric unit and P6-d.2, a C₃ symmetric protein comprising only helix D. The two protein structures are nearly identical with a backbone rmsd = 0.45 Å. The backbone structures of P6-d are in excellent agreement with the computational template [Cα rmsd = 0.40 Å (P6-d.2) and 0.61 Å (P6-d.1)]. Close agreement is also observed when a pair of dimers comprising adjacent proteins in the crystal are aligned and compared to the computational template. A pair of laterally neighboring proteins (P6-d.1—P6-d.2) has backbone rmsd = 0.64 Å, and an axially neighboring pair (P6-d.1—P6-d.1) has backbone rmsd = 0.70 Å (Fig. 3B). A comparison of peptides within the asymmetric unit to their coordinates in the model yields rmsd = 1.3 Å (all resolvable atoms) and rmsd = 0.68 Å (backbone only) (Fig. 5A).

Structural Characterization. Although the energy landscape strategy allows many possible interprotein separations and contacting mediating residues, in the P6-d crystal, backbone van der Waals contacts are present at the intralayer point of closest approach between proteins (6.4 Å between alpha helical axes). At this interface, the design yields a parallel GX₇G motif, often observed at interhelical contacts (45, 46, 50, 51). As discussed above, P6-a has a related antiparallel GX₇G₃A motif, which may be expected because the single template crystalline structure was iden-
identified using a model having a glycine exterior. P6-b, P6-c, and P6-e do not have this motif. Glycine and small amino acids have been suggested to promote crystallization through the reduction of surface entropy (8, 14, 52–54), and the presence of GX₃G motifs at each protein-protein interface is consistent with a tightly packed crystal. Furthermore the volume per molecular weight (Matthew’s coefficient) \( V_m = 1.99 \text{ Å}^3/\text{Da} \) is considerably lower than that typically observed for protein crystals \((V_m) = 2.68 \text{ Å}^3/\text{Da}, 1.5 \text{ Å}^3/\text{Da} < V_m < 5 \text{ Å}^3/\text{Da}) \) (21). The solvent content (36%) is less than that observed for typical protein crystals (51%), hexagonal protein crystals (57%), and protein crystals with four polypeptide chains in the asymmetric unit (51%) (21, 55, 56).

Backbone and side-chain interactions observed in the crystal structure are consistent with the parallel protein-protein orientation (Fig. 5 B and C). The C-terminal location of the GX₃G motif in P6-d imposes asymmetry along the helix favoring parallel orientations, whereas in P6-a, this motif is near the midpoint of the protein \( \left(G_0X_1G_1\right)_n \), and approximately 180° rotation about this point of contact permits antiparallel packing (Fig. 2 A and B).

The large residue (Tyr1) at the N terminus of P6-a may also contribute to destabilization of the parallel configuration. Charged side-chains often have high conformational entropy, and they are believed to disrupt crystal-packing interfaces (8, 52, 53, 57, 58). These residues are frequently mutated to induce crystallographic order (54, 59). Lys, in particular, appears infrequently at such interfaces (52, 53). Nevertheless, the P6-d structure provides evidence that ionizable residues such as Lys, Arg and Asp can form well-defined, complementary interactions in a designed, high-density crystal (Fig. 5 B).

The P6d structure suggests that well-defined crystal contacts likely require low conformational entropy amino acids to drive crystallization, e.g. GX₃G motif, augmented by complementary side-chain interactions (54, 59). Compared to P6-d, the models of P6-b, P6-c and P6-e have larger interprotein separations, \( R = 19.9, 20.2 \text{ Å}, \) and \( 19.1 \text{ Å} \), respectively, and do not contain the “small-X,-small” motif. Larger residues may be incorporated at the crystal contact positions, which may decrease the propensity toward crystalline ordering. In the model of P6-e, the point of closest approach is at the C terminus and the complementary backbone-backbone contact area is necessarily reduced compared to P6-d.

Conclusion. P6-d forms the first de novo designed protein crystal. The high-symmetry, high-density P6 structure possesses an uncommon polar arrangement throughout (6, 9), providing a route to controlled supramolecular parallel alignment of proteins (1, 6). De novo protein crystal design presents unique challenges. The high degree of self-assembly required to achieve a targeted crystal structure can also lead to aggregation and poor solubility. The probabilistic computational protein design methodology, however, provides a unique view of the sequence-structure energy landscape compatible with a chosen crystal lattice. Candidate
sequences and structures, which need not be nearby in either structure or sequence space, can be identified using an “aerial view” of the sequence-structure energy landscape, increasing the likelihood of identifying a protein that forms the targeted crystal structure. Computational protein crystal design can be of great utility to structural biology and genomics, where protein crystalization is essential to obtain high-resolution structures (10, 11). Partial (low-resolution) structural information or comparative models may be used in building model crystals and designing crystal contacts, potentially resulting in sequences with improved crystal quality and higher resolution structural information. Efforts in crystal design can further our understanding of the effects of mutation and modifying crystallization conditions. Targeting high-symmetry space groups and reducing crystal solvent content can improve the quality of X-ray diffraction data and simplify structure determination. Given the wide array of functionalities and cofactors that can be incorporated into proteins, targeted protein crystal design can also provide a vehicle to explore new protein-based materials and nanostructures.

Materials and Methods

Computational Design. The structure of the protein and the crystalline configurations were generated by applying the appropriate symmetry operations (translations and rotations) consistent with the C₃ symmetry of the inorganic and the P6 symmetry group. R is related to the unit cell parameters a = b = √3R, and the distance between the centers of mass of two interfac- ing proteins in adjacent layers is dictated by the c unit cell parameter. The rotational orientation between two neighboring interalayer proteins (θ) is defined in the plane perpendicular to the C₃ axis of a protein. Due to the protein’s C₃ symmetry, only the range 0° ≤ θ ≤ 120° is unique. The angle θ is varied by rotating the proteins about their C₃ axes such that the C₃ symmetry axis of adjacent proteins (and P6 symmetry) was retained. Variation of R and θ yields an ensemble of candidate crystalline structures. Structures with overlapping atoms or other high-energy interactions were filtered using AMBER (41) or CHARMM27 (60, 61) potentials, e.g., structures with R < 180 Å possess overlapping backbone atoms. In arriving at the sequence-structure energy landscape (Fig. 3A), a statistical thermodynamic method for calculating site-specific probabilities of the amino acids and the average energy over sequences was applied (20, 31). The method uses the AMBER energy function (41) and a discrete rotamer library (62). Iterative calculations, wherein the most probable amino acids are specified after each iteration, are used to identify specific sequences within minima on the sequence energy landscape.

Crystal Screening and Structural Determination. For initial crystal screening, peptides, P6-b, P6-d, and P6-e were ordered from Genscript (60 mg scale, desalt purity). To obtain larger quantities, sequences P6-a, P6-c, P6-d, and P6-e were synthesized (100 μmol scale) via solid-phase peptide synthesis, using 9-fluorenylmethoxycarbonyl (Fmoc), chemistry and upon resin clea- vage purified by reverse-phase high-performance liquid chromatography (HPLC). Crystals of P6-a and P6-d were grown at room temperature using the hanging drop vapor diffusion method and flash frozen in liquid nitro- gen prior to data collection. Multiple crystals of peptide P6-a were obtained with a peptide concentration of 10 mg/mL with reservoir solution (0.1 M cobalt (II) chloride hexahydrate, 0.1 M MES monohydrate pH 6.5, 1.8 M ammonium sulfate). A single crystal of peptide P6-d was obtained with a peptide concentration of 7.2 mg/mL with reservoir solution (0.17 M ammonium acet- ate, 0.085M tri-sodium citrate dihydrate pH 5.6, 25.5% vol/vol PEG 4000, 15% vol/vol glycerol). X-ray diffraction data were collected using a Rigaku R-Axis IV image plate detector equipped with a Cu Kα radiation source, and the structures were solved by molecular replacement. For protein P6-a, a poly-alanine model generated from coordinates of a single helix from a similar protein (PDB ID code 1COI) (38), was used as a search model. The re- finement strategy converged the Rwork/Rfree = 0.175/0.216. For protein P6-d, four helices from the computational design model served as the initial search ensemble and were used concurrently. The refinement strategy converged Rwork/Rfree = 0.148/0.205.

ACKNOWLEDGMENTS. We thank Axel Kohlmeyer for assistance with model- ling (Topo Tools); David Christianson, Dan Dowling, Julie Aaron, Pat Lombardi, Mustafa Koskal, and Steven E. Stayrook for their assistance with structure determination and refinement; E. James Peterson, Ivan Dimchowick, Jacob Goldberg and Glen Litszczak for their assistance with instrumentation and synthesis; and Ian Samihl for helpful comments. The authors acknowledge support from the Penn NanoBio Interface Center (National Science Founda- tion NSEC DMR-0425780), the US Department of Energy (DE-FG02- 04ER46156), the National Institutes of Health (HL085303, GM54816), and the University of Pennsylvania’s Laboratory for Research on the Structure of Matter (NSF MRSEC DMR05-20020) for infrastructural support.


